Massachusetts Institute of Technology
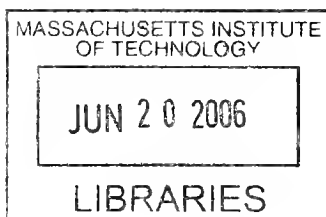Department of Economics
Working Paper Series

INTERNET RETAIL DEMAND: TAXES, GEOGRAPHY, AND ONLINE-OFFLINE COMPETITION

Glenn Ellison
Sara Ellison

Working Paper 06-14
April 28, 2006

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Internet Retail Demand:
Taxes, Geography, and Online-Offline Competition[1]

Glenn Ellison
MIT and NBER

and

Sara Fisher Ellison
MIT

April 2006

Abstract

Data on sales of memory modules are used to explore several aspects of e-retail demand. There is a strong relationship between e-retail sales to a given state and sales tax rates that apply to purchases from offline retailers. This suggests that there is substantial substitution between online and offline retail, and tax avoidance may be an important contributor to e-retail activity. Geography matters in two ways: we find some evidence that consumers prefer purchasing from firms in nearby states to benefit from faster shipping times as well as evidence of a separate preference for buying from in-state firms. Consumers appear fairly rational in some ways, but boundedly rational in others.

.

1 Introduction

The recent growth of Internet retail (e-retail) has attracted a great deal of attention in the academic literature and popular press.[1] While we have learned a lot about the structure of these markets the past few years, the remarkable swings in the market value of e-retail companies provides ample evidence that our understanding of the industry's future is quite limited.[2] This future of e-retail is of interest for both intellectual and practical reasons. Intellectually, e-retail is a great case study: it provides an opportunity to reexamine our understanding of consumer and firm behavior and suggests new questions. Practically, e-retail could have significant effects on the economy. It will not remain small for long at its current growth rate: it has grown steadily at about 25% per year since the collapse of the dot-com "bubble." And even a small e-retail industry could have a substantial impact on traditional retail, which employs as many Americans all manufacturing industries combined.[3]

In this paper we investigate aspects of consumer behavior that will have a substantial impact on the future of Internet and traditional retail. We focus on three issues. First, we examine substitution betweeen Internet and traditional retail. It is not clear whether e-retail or traditional retail will prove more efficient in the long run. Whether the two are close substitutes, though, will have a dramatic impact on the future of the less efficient channel. Second, we examine the extent to which the success e-retail has had is due to the *de facto* tax-free status of most e-retail purchases in the U.S.[4] This bears on the relative efficiency of e-retail, and is important to understanding what may happen if states are able to tax online sales, with the Internet Tax Nondiscrimination Act set to expire in

---

[1] See, for example, Goolsbee (2000), Smith (2001), Chevalier and Goolsbee (2003), and Ellison and Ellison (2005).

[2] Amazon's market value, for example, grew from $2 billion at the start of 1998 to $35 billion in the middle of 1999, and fell back to $3 billion in 2001 before reaching $25 billion for a second time in 2003.

[3] U.S. e-retail sales are now approximately $100 billion per year, which is 3% of total retail sales.

[4] Forty five U.S. states levy sales taxes on traditional retail purchases. Each of these states also has laws assessing "use taxes" on purchases that its residents make from out-of-state firms. However, the Supreme Court ruled in *Quill vs. North Dakota* (1992) that absent new federal law, a state could not compel a firm without substantial physical "nexus" in that state to collect use taxes on its behalf. The 1998 Internet Tax Nondiscrimination Act makes explicit that web presence alone does not constitute nexus. While consumers are obligated to self-report use-tax liability, few do in practice. Note that states are able to collect sales taxes on e-retailers' in-state sales.

2007.[5] Third, we examine the geography of e-retail. It is commonly supposed that geographic differentiation is an important factor allowing traditional retail stores to maintain the markups over marginal cost they need to survive. Branding, obfuscation, or other factors may allow e-retailers to survive even without geographic differentiation, but knowing whether geographic differentiation is really eliminated is also important for understanding what market structure might evolve.[6]

Although everything we study is in the context of e-retail, this paper can also be thought of more generally as an empirical analysis of consumer informedness and sophistication. The standard fully-informed, rational analysis would make a number of simple predictions, e.g. consumers should compare firms on the basis of the tax-inclusive prices and make decisions on the basis of the prices charged at the time when the purchase is made (as opposed to at other times). Alternate predictions could be obtained in a number of ways. For instance, one could use rational models with information acquisition costs, rational models with information processing costs, or models with "irrational" consumers. We find that several of the standard predictions appear not to hold in our data and discuss what this suggests about the form of our consumers' bounded rationality.[7]

The environment we study is that examined in Ellison and Ellison (2004): we look at consumers shopping for computer memory modules using the Pricewatch.com search engine. For a period of approximately one year, we have hourly data on the twelve lowest prices listed on Pricewatch for each of several products. We know the state in which the e-retailer listing each price is located. Our quantity data is unusually good in one respect and unusually bad in another. The bad part is that we only observe purchases from two of the listed websites (both located in California), so we do not know how many consumers purchased from other websites (or from traditional retailers). The good part is that the

---

[5]There are two ways in which the *de facto* tax-free status of Internet purchases in the U.S. might be threatened in the near future. The expiration of the Internet Tax Nondiscrimination Act in 2007 could have implications for the legal definition of nexus. In addition, eighteen states have joined the Streamlined Sales Tax Project in an attempt to simplify and harmonize their sales tax laws. The Project's goals are to encourage online retailers to agree to collect use taxes for sales made in those eighteen states and, eventually, to pave the way to federal legislation requiring collection of use taxes.

[6]See Brynjolfsson and Smith (2001), Chevalier and Goolsbee (2003), Baye and Morgan (2004), Ellison (2005), and Ellison and Ellison (2004).

[7]Hossain and Morgan (2006)'s analysis of bidding on eBay provides compelling evidence of limitations in consumer rationality in that environment.

data are at the individual order level and include each consumer's location.

The structure of the data provides several nice opportunities for examining consumer preferences and behavior. First, the fact that we observe the state in which each consumer is located creates an opportunity to look at geography, taxes, and online-offline substitution: we can quantify the extent to which our websites sell more in states that levy higher sales taxes—taxes primarily affect the firm's competitive position relative to traditional retailers—and to states that are nearby. Second, there is substantial turnover in the Pricewatch lists, both in terms of which websites make the list of the twelve lowest-priced and in their price ranking. Hence, there are many hours in which our two websites are mostly competing against other California e-retailers, and others in which they are competing against e-retailers in New Jersey, Illinois, Oregon, etc. with similar prices. Looking at how state-specific sales in a given hour are affected by the competitors' locations is another way to identify geography and tax effects. Third, wholesale prices for memory modules are remarkably volatile. In the highly competitive Pricewatch universe, wholesale price increases and decreases are passed through very quickly. Many traditional retailers, in contrast, keep prices fixed over the course of each week at the price they advertised in the latest Sunday sales circular. This creates a interesting source of variation between online and offline prices: in some weeks the online-offline price gap is much lower on Friday than it was on Sunday, and in other weeks it is much higher. How consumers react to this price gap will again be informative on online-offline substitution and consumer awareness of up-to-date price information.

The paper is organized around two analyses designed to exploit different sources of variation. In Section 3 we exploit the time-invariant factors—state-level tax rates and differences in state-to-state shipping times—in the simplest way possible. We run cross-section regressions examining the total number of orders received from each state over the course of the year. These regressions provide clear evidence that tax savings are an important motivation for online shopping: our e-retailer's sales are substantially greater in high-tax states than in low-tax states. We can provide an additional piece of supporting evidence to bolster the case that the differences are due to taxes and not due to unobserved consumer heterogeneity: our e-retailer sells *much* less in California than in comparable

3

states. (This would be expected under the tax hypothesis because our e-retailer must charge sales tax on sales to California residents.) These cross-section regressions provide some weak evidence that geography matters for shipping-time reasons and that consumers are somewhat more sensitve to differences in tax rates for more expensive products (for which the tax difference in dollars is larger).

Section 4 applies standard demand estimation techniques in an unusual way to exploit the hourly variation in the data: we estimate discrete choice models that use as their dependent variable the number of orders of a given product from consumers in a particular state in a particular hour.[8] The nonstandard part of the application is that we only have data on consumer purchases from two of the listed firms. Normally, one applies discrete choice models to datasets containing all firms' market shares. Having data on all firms is, however, not necessary to identify the model given that we have substantial intertemporal variation in the characteristics of the competitors. It is this variation that helps us learn about substitution between e-retailers, how much attention consumers pay to geography, taxes, and so forth, simply by looking at how our firm's sales go up and down as rivals' prices and locations change.

The discrete-choice analysis provides some evidence that consumers pay attention to differences in the taxes between e-retailers. There is little evidence of geographic differences between retailers due to differences in shipping times. Geography does appear to have one significant effect: consumers are estimated to have a strong preference for purchasing from e-retailers located in their own state (after controlling for differences in shipping times and sales taxes).

A general theme that emerges from the discrete-choice analysis is that consumers seem far from the fully-informed, fully-rational ideal. Consumers react very strongly to price differences in settings where the price comparisons are easy, such as between competing e-retailers. They react less strongly to tax differences of a similar magnitude. They react less strongly still, hardly at all, to transitory variation between online and offline prices of a similar magnitude. These findings are roughly consistent with models of costly information

---

[8]These regressions include dummy variables for each state so that the results derive from variation that is independent of the variation that identifies the cross-section regressions of Section 3.

4

acquisition or rule-of-thumb reasoning.

Our work is related to a number of previous papers. The standard reference on Internet taxation is Goolsbee (2000). It examines a 1997 survey in which 25,000 consumers were asked (among many other things) whether they had ever bought products online. Consumers living in states with higher sales tax rates are found to be more likely to have bought products online. The big-picture conclusion is that subjecting e-retailers to taxation could reduce online sales by 24%. One motivation for the the tax part of our paper is to address a couple potential concerns about Goolsbee's work: an elasticity derived from analyzing whether consumers ever purchase anything on the Internet could be very different from the elasticity of total quantity with respect to taxes (which will reflect much more the behavior of intensive Internet shoppers); and one could also worry that some of the tax effects he finds could be due to differences in unobserved consumer characteristics across states (driven, for example, by California and Washington having high sales taxes as well as populations inclined to use the Internet).[9] Our tax results also relate, of course, to the literature on the effects of sales taxes on location and consumer behavior in traditional retail, e.g. Fox (1986) and Walsh and Jones (1988).

A number of other papers have used data from price search engines to examine aspects of e-retail demand. Brynjolfsson and Smith (2001) examines consumers who visited EvenBetter.com in 1999. It has a puzzling finding on taxes: consumers are estimated to be twice as sensitive to differences in taxes as they are to differences in item prices.[10] It also finds strong evidence that consumers prefer branded e-retailers over lesser known firms. One limitation is that they do not actually have any quantity data. The quantity data is imputed by assuming that that consumers purchased from the e-retailer they visited last. Ellison and Ellison (2004) examines the same Pricewatch data as this paper. It notes that websites attracting customers via Pricewatch.com have extremely price-elastic demand, and

---

[9]Despite the examples of California and Washington, sales taxes in the U.S. are, in fact, not positively correlated with the demographic controls for computer usage we employ. For example, Louisiana, Tennessee, Oklahoma, and Alabama each have both one of the eight highest average tax rates in the country and a below average fraction of households with home Internet access. Goolsbee casts doubt on the unobserved heterogeneity explanation for his results by using extensive household-level demographic controls, by including MSA dummies, and by showing that tax rates are not correlated with ownership of computers.

[10]This could be explained as an artifact of price endogeneity if higher prices are associated with higher unobserved quality whereas higher taxes are not.

investigates how it is that firms are able to maintain nontrivial markups. The primary observations on this count are that firms engage in a great deal of obfuscation, and that an adverse selection disincentive for price cutting, like that described in Ellison (2005), appears to be present. Baye, Gatti, Kattumen and Morgan (2005) examine clickstream data on consumers shopping for PDAs through the Kelkoo.com search engine in 2003. They find price sensitivity and take advantage of the structure of their data to address a number of other questions: how price-sensitivity varies with the number of listed firms; how screen- and price-rank separately influence demand; etc.

Several papers have addressed online-offline competition with limited data. Brown and Goolsbee (2001) find that in the mid 1990's term life insurance rates dropped more for demographic groups whose members were more likely to have Internet access. Goolsbee (2001) constructs a measure of the competitiveness of local retail markets using survey data on the prices paid for consumers and shows that consumers in less competitive traditional retail markets are more likely to buy computers online.[11] Prince (2005) also examines online and offline substitutability of personal computer purchases using the same measure of competitiveness of traditional retail markets. Chiou (2005) examines consumer's decisions on where to purchase DVDs using a dataset that includes both online and offline purchases.

We are not aware of any other work on spatial differentiation between e-retailers. A number of papers have examined spatial differentiation in traditional retail, including Weisbrod, Parcells and Kern (1984), Chiou (2005), and Davis (2006).

2   Data

In this paper we examine sales of four different types of memory modules, 128MB PC100, 128MB PC133, 256MB PC100, and 256MB PC133.[12] Our price data were obtained by downloading the first (or first and second) screens from Pricewatch's memory price lists

---

[11] These results could in part reflect unobserved heterogeneity: the high prices paid for computers in an area could also reflect the presence of consumers who are more computer savvy and purchase computers that are of higher quality in the unobserved dimensions.

[12] As described in Ellison and Ellison (2004), our e-retailer sells three versions of each of these types of memory modules. The three versions are clearly ranked in quality. In this paper, we restrict our attention to the lowest quality "generic" version of each type of memory module. This is the only quality level for which one can easily use Pricewatch to identify competitors' prices. Low quality memory also accounts for the majority of our firm's sales.

on an hourly basis from from May 2000 to May 2001 (with some gaps). Our data on the 128MB modules include information on the twenty four lowest-priced websites listed on Pricewatch. The data on 256MB modules include information on the twelve lowest-priced websites. There is a fair amount of turnover and reshuffling of the price lists from day to day (and even from hour to hour in some periods). Over the course of the year there is a dramatic decrease in prices. For example, in the space of a year the price of a 128MB modules fell from about \$120 to about \$20.

Pricewatch does not calculate sales taxes for consumers on these pages, but it does list the home state of each retailer so that a consumer who knew the tax rate in his home state (and understood that sales taxes will apply if and only if he or she buys from an in-state firm) could take sales tax differences into account. We downloaded the state locations as well.

We obtained quantity data for these products from an Internet retailer that gets most of its traffic from Pricewatch. It operates two similar websites, which typically have different prices for the products studied.[13] The quantity data again cover May 2000 to May 2001 with some gaps. The raw data are at the level of the individual order. We use data on approximately 15,000 orders. The available data on each order include the website from which the customer made the order, detail on what was ordered, and the shipping address. Our e-retailer is just one of many listing products for sale on Pricewatch. A rough estimate is that 100,000 other consumers visited Pricewatch during our sample period and purchased a corresponding product from one of the e-retailers for which we do not have quantity data.

We also use a few state-level variables. The most important of these is the state's average sales tax rate. Sales tax rates vary by county and locality in many states. Our data are averages across the various jurisdictions within a state computed by a private firm. We collected data on UPS ground shipping times by querying the UPS website. These data include both shipping times from our e-retailer's zip code to each state, and a state-to-state shipping time matrix.[14] Our other state level variables come from Census Bureau datasets:

---

[13]There are several possible motivations for having multiple websites: they may be given different looks and consumers may have heterogeneous reactions; it allows the websites to be more specialized (which seems to be attractive to some consumers); it facilitates experimentation; it may help promote private-label branded products; The firm may occupy multiple places on the Pricewatch screen.

[14]UPS provides these data on a zip code to zip code basis and there can be some within-state variation.

the fraction of households with home Internet access as reported in a 2001 survey, the population of each state in the 2000 census, and the number of computer stores and gas stations reported in the 1997 Census of Retail Industries.

3   Analysis of aggregate state-level sales

In this section we take the most straightforward approach to examine how the time invariant variables in our dataset—sales tax rates and shipping times—affect consumer demand. We construct measures of the total number of orders received from each state, and use regressions to, for example, look at whether our e-retailer sells more in states with high sales taxes than in states with low sales taxes.

3.1   Summary statistics

The regressions in this section will have 51 observations: one for each state and the District of Columbia. We use two primary dependent variables: *Quantity128* is the number of orders for 128MB modules received over the course of the year from a given state; *Quantity256* is the corresponding number for 256MB modules.[15] Summary statistics for the basic regressions are presented in Table 1. Our e-retailer sells 204 128MB memory modules to the average state over the course of the year. This ranges from a low of 19 in the District of Columbia to a high of 762 in Texas. Unit sales of 256MB memory modules are about half as large. The average sales tax rate is 5.7 percent. Four states have no state or local sales taxes. The UPS ground shipping time from our retailer to the average state is about 4 days.[16] The percentage of households with home Internet access varies from a low of 40.6% in the District of Columbia to a high of 70.2% in New Hampshire. The average state has 230 computer stores. The ratio of computer stores to gas stations ranges from a low of 0.041 in West Virginia to a high of 0.184 in California.

We typically collected data using one zip code from the the largest population center in the state. In some cases where a state did not have one dominant population center and the shipping time varied we took an average of the times for the two largest population centers.

[15]Note that in doing this we are summing both over the two websites for which we have data and over the two speeds of each size memory module: PC100 and PC133. We do this because there is no reason to expect that taxes or geography would have a different impact across websites or speeds.

[16]The minimum value of 1.5 days reflects that shipping times are one day for shipments to Southern California and two days for shipments to Northern California.

Although prices are not used in this state-level anaysis, they are relevant for the interpretation of some results. The mean price of a 128MB memory module is \$70. The mean price of a 256MB memory module is \$139. A one percentage point difference in tax rates, then, adds 70 cents on average to a 128MB module but \$1.39 to a 256MB module.

3.2   Basic results

To analyze how the number of orders received from state s is related to the state's tax rate we estimate a negative binomial regression model, i.e. we assume

$$
\begin{aligned}
Quantity_s \quad &\sim \quad \text{Poisson}(\mu_s) \\
\log(\mu_s) \quad &= \quad \beta_0 + \beta_0 + \beta_1 \, OfflineSalesTaxRate_s + \beta_2 California_s + \beta_3 ShippingTime_s \\
&\quad + \beta_4 \frac{ComputerStores_s}{GasStations_s} + \beta_5 InternetAccess_s + \beta_6 \log(Population_s) + \epsilon_s,
\end{aligned}
$$

where the ϵ_s are independent random variables with $e^{\epsilon_s} \sim \Gamma(\theta, \theta)$, and estimate the parameters by maximum likelihood.[17] One can think of this as similar to estimating a linear regression with $\log Q_s$ as the dependent variable.[18]

Table 2 presents coefficients obtained from estimating the regression above on the total unit sales to each of the 51 states. The first column uses 128MB memory module sales as the dependent variable. The results are strongly suggestive that sales taxes have a large effect on online sales. The 5.94 coefficient estimate on *OfflineSalesTaxRate* indicates that a one percentage point increase in a state's sales tax increases the number of orders our e-retailer receives from that state by about 6%. The average sales tax rate in our data is 5.7%. Hence, in a typical state, online purchases would be predicted to decrease by about 30% if the offline sales tax were eliminated. Goolsbee argues that this is a good forecast

---

[17]The Poisson regression model is the special case of the negative binomial with $\theta = \infty$. In applied work it is common to find that a specification test can reject the Poisson model in favor of other models that allow for more dispersion. The particular assumption that the errors are distributed like the logarithm of a gamma random variable (as opposed to being normally distributed for example) is motivated by the fact that a relationship between Poisson and gamma random variables allows the likelihood to be evaluated without a numerical integration. The distribution of Q_s turns out to be negative binomial which is what gives the model its name. Section 19.9.4 of Greene (1997) provides a clear description of the model. Hausman, Hall and Griliches (1984) discuss a number of models for count data.

[18]The advantage of the negative binomial regression is that the model can be estimated in the same way regardless of whether some of quantities are zero. All quantities are positive in our base analysis, but there will be some zeros in later analyses of quantities sold during particular time periods. The third and fourth columns of table 2 show coefficient estimates from regressions with $\log(Q_s)$ as a dependent variable for comparison. The results are similar.

for the impact of taxing online sales—the implicit assumption is that achieving tax parity between online and offline retail should have a similar effect regardless of whether it is achieved by increasing online taxes or by decreasing offline taxes.

The extra customers our firm attracts in states with high sales tax rates could in priniciple be coming from three sources: they might otherwise have purchased from traditional retailers, from other online retailers, or not at all. Few of our e-retailer's online competitors are in any particular state (other than California), so very little of the added demand could be taken from online retailers in the customer's home state. It also seems unlikely that most could be people who otherwise would have chosen not to buy memory, because this would require a highly elastic aggregate demand for memory.[19] We conclude that there must be substantial substitution between online retail and offline retail.

The coefficient on the *California* dummy provides additional support for the view that what we have estimated is a tax effect and not an artifact of unobserved state-level heterogeneity. What would we predict about our firm's sales to California if the coefficient on *OfflineSalesTaxRate* is truly a tax effect? First, since our firm has no tax advantage relative to brick and mortar stores in California—its California customers must pay sales tax—we would expect its sales to be about 35% lower than one would otherwise predict given state covariates. (California sales are taxed at 7.25%.) Second, our firm has a disadvantage relative to non-California e-retailers when selling in California. One would expect that this disadvantage would lead to an additional reduction in sales. The estimated coefficient on *California* indicates that sales to California customers are about 67% lower than sales to comparable states. It is implausible that an effect of this magnitude could be due to an unobserved distaste for online shopping on the part of Californians.

The estimate on the *ShippingTime* variable provides some weak evidence that geography still matters on the Internet. Sales are estimated to be reduced by about 10% if UPS ground shipping to the destination state is one day longer.

The coefficients on the other control variables seem reasonable. Sales are higher in states where the fraction of residents with Internet access is higher. We cannot reject that

---

[19]The regressions of quantity on the log of the lowest listed price reported in Ellison and Ellison (2004) suggest that the elasticity of aggregate demand with respect to price may be close to one.

the coefficient is one, which would correspond with sales being proportional to the number of people with home Internet access. The coefficient on the computer store-gas station ratio might be expected to have either sign: it reflects both interest in computers and the availability of computer parts at traditional retail stores. The estimated coefficient is positive but not statistically significant. Population is obviously a strong determinant of aggregate sales. Potential reasons why the coefficient might be less than one include that population is an imperfect proxy for the potential market size (which is affected by income, business activity, and other factors), and that larger population states may have better offline retail.

The second column of Table 2 presents coefficient estimates from a regression with orders for 256MB memory modules as the dependent variable. These results are very similar: sales are substantially higher in states that levy higher sales taxes on traditional retail purchases; sales are notably lower in California; there is weak evidence that shipping times may affect sales; the effects of the other demographic variables are similar.

The third and fourth columns of Table 2 report demand estimates obtained via OLS regressions with $\log(Quantity)_s$ as the depedent variable. The results are quite similar to those from the negative binomial regressions.

3.3   Demand at different price levels

The fact that we have data on goods sold at different prices provides an additional opportunity to gain insights into consumer behavior: we can compare the effects of sales taxes on sales of expensive and inexpensive products. What would we expect to find in such a comparison?

A good way to think about this is in terms of a discrete-choice model with heterogeneous preferences for online vs. offline shopping: suppose a consumer of type θ in market i gets utility $v_{i,on} - p_{i,on}$ if she buys online and $v_{i,off} - p_{i,off}(1 + t) + \theta$ if she buys offline and that the CDF of θ in market i is F_i. In such a model, a dt increase in the offline tax rate increases the effective offline price by $p_{i,off}dt$ and thereby increases online demand by $p_{i,off}f_i(p_{i,off})dt$. The answer is therefore that estimated coefficients on the tax rate may differ across products, and whether they do should reflect how the distribution of consumer

11

preferences for offline vs. online shopping compares for the two products, expensive and inexpensive. How we would expect these distributions to compare depends on where the θs come from. For example, if θ was primarily determined by a consumer's taste for computer use vs. driving, then the the most natural assumption would be that the distribution of θ would not vary across products. In this case, the sensitivity of online demand to the offline tax rate should be proportional to the product price. Alternatively, the distribution of θ could arise from heterogeneity in consumers' willingness-to-pay to get the product immediately. In such a case, it would be more natural to assume that what is invariant across products is the percentage of the product price that a consumer is willing to pay to avoid waiting for the product to arrive. In this case, the density of θ would be inversely proportional to the product price, and the sensitivity of online demand to the offline tax rate would be constant across products.

Our dataset has two sources of variation that let us make such comparisons. First, we can compare the demand for 128MB modules with the demand for 256MB modules. The sales-weighted mean price of a 256MB memory module is about 60% higher than that of a 128MB module. Looking at the first and second columns of Table 2 we note that the coefficient estimate on *OfflineSalesTaxRate* for 256MB modules is slightly larger than that for 128MB modules. The standard errors, however, are sufficiently large so that we can reject neither that they are equal nor that they differ by 60%.

Second, we can exploit the substantial time-series variation in prices by comparing demand in different time periods. For example, we look at the demand for 128MB modules in the time period when they were over $100 compared with a period a few months later when they were much cheaper. The first two columns of Table 3 contain such a comparison.[20] The estimates in the first two columns indicate that quantities were more sensitive to differences in tax rates in the period when 128MB modules were more expensive, suggestive that consumers' channel preferences are more likely fixed and not dependent on item price. Again, though, the estimates are not sufficiently precise to allow us to reject equality. The

---

[20]The former period is at the end of our data, whereas the latter is mostly in the summer of 2000. Each of these is again estimated via a negative binomial regression run on a cross-section containing 51 observations. By "obtained from two separate time periods" we mean that the first dependent variable is obtained by summing the hourly sales to each state over the set of hours during which the lowest price on Pricewatch was between 20 and 50 dollars.

evidence in the third and fourth column is less conclusive. The point estimate is about 20% larger in the high-price period, but the differences are far from significant.

The results on the *California* dummy are similar. In Table 3 the coefficient on the *California* dummy is larger in the period when the products are more expensive, but the differences are not significant. In Table 2 the point estimates are not larger for the more expensive product.

In summary, the comparisons across products and across times provide quite limited evidence of demand being more sensitive to differences in tax rates for more expensive items. It could be that the effect is there and data limitations prevent us from seeing it. Alternatively, it could be that the tax-sensitivity does not vary much. Such a result could be explained in two ways. First, it is possible that channel preferences may be scaling up with the item price, as mentioned above. A primary source of online-offline differentiation could be heterogeneity in the disutility for waiting for online purchases to arrive. A second "irrational" explanation would be that consumers may follow fairly simple rules of thumb. For example, the ones in high tax states may have learned that it is generally a good idea to buy products online and save on the sales tax, but have not developed more sophisticated rules recognizing that the tax savings are larger on more expensive items.

4   A Discrete-Choice Analysis

The Pricewatch environment exhibits an unusual degree of short-term variation in competitive conditions. This variation provides a nice opportunity to gain additional insight into e-retail demand and consumer behavior. In this section we use discrete-choice models to explore substitution between online and offline retail, substitution between e-retailers, the effects of geography and sales taxes, and consumer sophistication.

4.1   Motivation

The analysis in this section is designed to exploit two sources of short-term variation in our data: turnover in the relative price rankings and changes in price levels that are common to most firms. We briefly discuss each of these to provide some intuition for what the variation is and why it should be useful.

4.1.1 Turnover in price rankings

Figure 1 is a particularly clean example of the type of turnover in the relative price rankings we see multiple times a day. As such, it provides a nice illustration of how we can exploit this turnover to estimate aspects of consumer behavior. It shows the twelve e-retailers listed on the first screen of Pricewatch's 128MB PC100 memory page at 9am and 11am on August 1, 2000.

Two of the e-retailers made price changes between these two times. Coast-to-Coast Memory of New Jersey, which offered the lowest price of $112 at 9am, raised its price sufficiently so as to disappear from the top twelve by 11am. UpgradePlanet.com of Virginia, which was on the second page of the 9am list at $128, reduced its price to $111 and took over the top slot. The first three columns show information presented on Pricewatch: the e-retailers' names, their locations, and their prices. The fourth through the sixth columns contain numbers not presented on the Pricewatch site but which consumers could compute from the given information: the tax-inclusive prices that customers in New Jersey, Virginia, and California, respectively, would pay if they purchased from each of the e-retailers.[21]

Recall that we observe sales for two websites. However, we observe not just total sales but sales *into each state* at each hour. This fact, along with the turnover in relative price rankings, is crucial for our estimation strategy. To illustrate, consider first the case where we had sales by every firm into every state every hour. One could assess whether consumers pay attention to tax differences by looking at whether UpgradePlanet was making many more sales into New Jersey at 11am than Coast-to-Coast was making at 9am, and at whether Coast-to-Coast was making more sales into Virginia at 9am then UpgradePlanet was making at 11am. A preference for buying from nearby firms could, of course, offset the tax disadvantage. Geographic preference, however, could be separately identified in many ways. One could look at whether UpgradePlanet sells less than Coast-to-Coast had in states bordering New Jersey and more in states bordering North Carolina. One could use data from other points in time to look at how market shares in Oregon change when an Oregon firm is or is not present at the top of the list. (Oregon has no sales tax.) One

---

[21] A consumer, of course, would need to know his or her local sales tax rate and the fact that sales taxes are only assessed on in-state sales to make this calculation.

Information on Pricewatch			Price	Price	Price
Website	State	Price	into NJ	into VA	into CA
Pricewatch ranking at 9:01am EDT					
Coast-to-Coast Memory	NJ	112	118.72	112	112
Connect Computers	CA	113	113	113	121.64
Computer Craft	FL	114	114	114	114
Advanced PCBoost	CA	115	115	115	123.80
1st Choice Memory	CA	116	116	116	124.87
Jazz Technology	CA	117	117	117	125.95
Memplus.com	CA	117	117	117	125.95
Portatech	CA	119	119	119	128.10
Augustus Technology	CA	120	120	120	129.18
EconoPC	IL	120	120	120	120
Advanced Vision	CA	121	121	121	130.26
Computer Super Sale	IL	122	122	122	122
Pricewatch ranking at 11:01am EDT					
UpgradePlanet.com	VA	111	111	115.99	111
Connect Computers	CA	113	113	113	121.64
Computer Craft	FL	114	114	114	114
Advanced PCBoost	CA	115	115	115	123.80
1st Choice Memory	CA	116	116	116	124.87
Jazz Technology	CA	117	117	117	125.95
Memplus.com	CA	117	117	117	125.95
Portatech	CA	119	119	119	128.10
Augustus Technology	CA	120	120	120	129.18
EconoPC	IL	120	120	120	120
Advanced Vision	CA	121	121	121	130.26
Computer Super Sale	IL	122	122	122	122

Figure 1: Sample Pricewatch rankings: 128MB PC100 memory modules on August 1, 2000

15

could also examine tax effects controlling for geographic preferences by looking at relative magnitudes: does the $6.72 tax that Coast-to-Coast must levy in New Jersey have a larger effect than the $4.99 tax that UpgradePlanet must levy in Virginia?

Our quantity data are more limited: we observe sales for only two websites rather than all websites. This limitation, however, should not prevent us from examining how consumers react to different competitive enviromnents: the reactions mentioned above would be reflected in changes in the demand for the websites for which we do have data.

To think about how this works, suppose that our sales data were from Connect Computers.[22] At 9am Connect Computers' tax-inclusive price for New Jersey residents is lower than that of any other website. At 11am Connect Comptuters' tax-inclusive price for New Jersey residents is only the second lowest. Accordingly, if consumers pay attention to sales taxes we would expect Connect Computers' sales into New Jersey to be higher at 9am than at 11am. Similarly, its sales into Virginia would be higher at 11am than at 9am. We can estimate tax effects controlling for a home state preference by looking at how the magnitude of the 9am-11am drop in Connect's New Jersey sales compares with the 9am-11am increase in Connect's Virginia sales. A comparison of Connect's California sales at 9am and 11am will teach us about substitution between retailers: shipping times from New Jersey and Virginia to California are the same, so the comparison should help us learn how many consumers shift from the second-lowest to the low-priced firm when the low-priced firm reduces its price by one dollar.

4.1.2 Within-week changes in price levels

The second useful source of variation in our data is within-week changes in the level of e-retail prices. Traditional retailers do not change their prices frequently. In particular, it is common to advertise prices for memory modules in weekly sales circulars, so that the prices remain constant each week from Sunday-Saturday. The prices listed on Pricewatch, in contrast, are highly volatile. Many e-retailers hold essentially no inventory and pass on wholesale price changes almost immediately.

Figure 2 provides an illustration. The thin line is the lowest price on Pricewatch for

---

[22] Connect Computers is, in fact, not one of the websites from which we have data.

a 128MB PC100 memory module.[23] There are many instances where the price changes substantially over the course of a week. Usually these are price decreases. For example, between Sunday, September 17, 2000 and Saturday, September 23, 2000, the price dropped from $89 to $78. If consumers are rational and fully informed (and online and offline retail are close substitutes), then one would expect that an online retailer would sell more on this Saturday than they had on the previous Sunday. Prices also rise at times. For example, between Monday, November 20, 2000 and Friday, November 24, 2000, the online price increased from $42 to $53. In such a week, one would expect that online retailers would do worse on Friday and Saturday than they had on Sunday and Monday. Our data are well-suited to examine such predicitons.

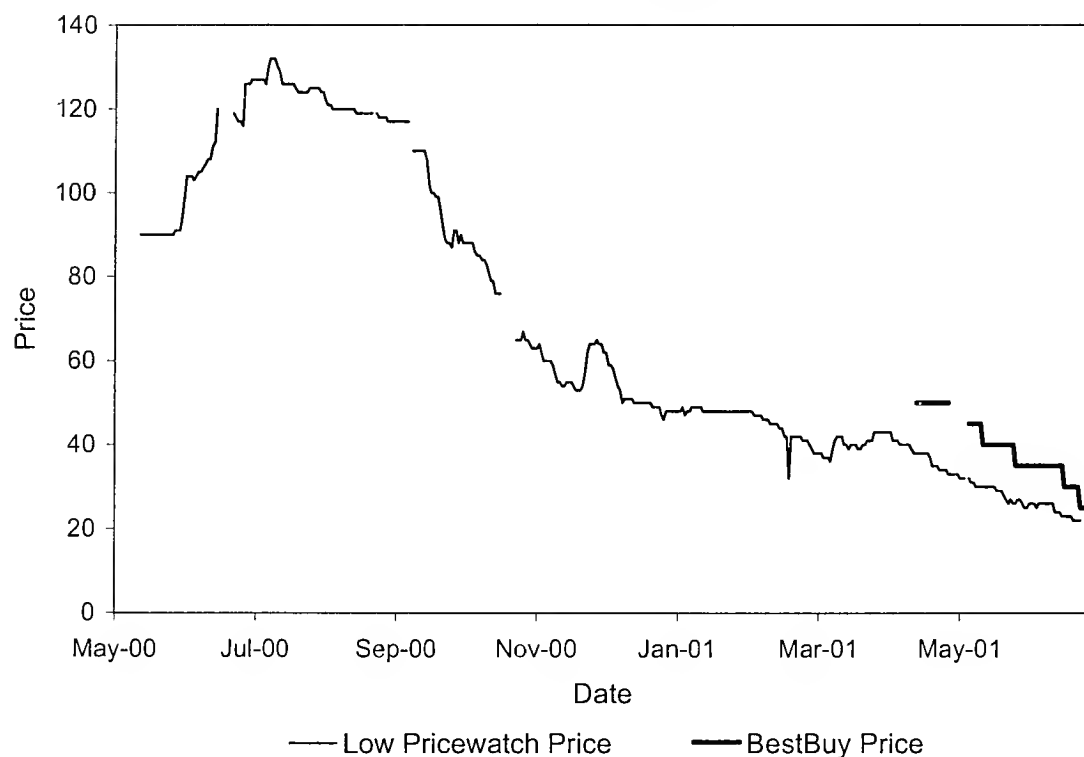**Online and Offline Prices: 128MB PC100 Memory**

Figure 2: Online and Offline Prices: 128MB PC100 memory modules

[23]To make the two price series more comparable, the shipping and handling fee that is standard on Pricewatch, $11, has been added to the listed price.

Toward the end of our sample period we collected price data for a comparable product from the largest traditional electronics retailer, BestBuy.[24] The bold line on the right side of the figure is a graph of these prices. If we had similar data for our entire sample period (and data from more retailers), we could try to estimate online-offline substitution by using the price gap as a primary explanatory variable. We do not have these data for most of our sample period, however, so we will just estimate the effect of within-week changes in online prices and employ a set of time trends to control for longer-term trends in the online-offline price gap.

4.2 Methodology

Let N_{sht} be the number of consumers in state s purchasing a particular type of memory module in hour h of day t from the twenty-four (or twelve for 256MB modules) websites whose prices we observe. Assume that consumer k's utility if he purchases from website i is

$$u_{iksht} = \beta_1(Price_{iht} + \beta_2 SalesTax_{isht}) + \beta_3 ShippingTime_{is}$$
$$\beta_4 HomeState_{is} + \beta_5 SecondScreen_{iht} + \epsilon_{ik},$$

where $SalesTax$ is the sales tax in dollars due on the purchase, $ShippingTime$ is the UPS ground shipping time, $HomeState$ is a dummy variable for whether website i is in state s, $SecondScreen$ is a dummy indicating whether website i only appears on the second screen of results, and ϵ_{ik} is a logit random variable independent of the right hand side variables (and of the additional right hand side variables and the error η_{hst} introduced below).

Writing X_{sht} for the vector of attributes on the right hand side of this expression, we have the familiar logit formula for the number of consumers in state s buying from website i conditional on the total number of purchases N_{sht}:

$$E(Q_{isht}|X_{sht}, N_{sht}) = N_{sht}\frac{e^{\beta X_{isht}}}{\sum_{j=1}^{24} e^{\beta X_{jsht}}}$$

Our dataset only contains sales from two particular websites. It does not contain the number of consumers purchasing from other websites, from traditional retailers, or not at

---

[24]The BestBuy product is a branded product that may be of higher quality than the products covered by the online price data.

all. The total number of consumers buying through Pricewatch is affected by a number of factors: there are clear day-of-week and hour-of-day effects; Internet use is climbing over our sample period; there are substantial price declines that should increase aggregate demand; there is variation in the online-offline price gap; and there may be intertemporal price effects with the size of the potential consumer pool at a given time being affected by past prices. Our data will not allow us to separately identify all of these effects. The approach we take is simply to specify a flexible functional form for the aggregate Pricewatch demand that could reflect each of the effects. Specifically, we assume

$$N_{sht} = \delta_s \bar{q}_h e^{\gamma_1 MinPrice_{ht} + \gamma_2 SundayPrice_{ht} + \gamma_3 Weekend_t + \gamma_4 TimeTrend1_t + ... + \gamma_7 TimeTrend4_t} + \eta_{hst},$$

where δ_s is a state fixed effect to be estimated, $\bar{q}_h$ is an hour-of-day fixed effect, $Minprice_{ht}$ is the lowest price listed on Pricewatch, $SundayPrice_{ht}$ is the price on the most recent Sunday, $Weekend_t$ is a weekend dummy, the $TimeTrend$ variables allow for linear time trends with slopes changing every ninety days, and η_{hst} is a random error term assumed to have mean zero conditional on the right hand side variables in this equation.[25]

We estimate the model via nonlinear least squares, using hour-website-destination state sales as the dependent variable. The model could in principal be estimated on the 800,000 observation datasets obtained by using sales to each of the fifty-one states in each of the approximately 7900 hours by each of the two websites as the observations. Some states, however, account for a very small portion of the sales in our dataset. Other states account for a nontrivial number of purchases, but rarely or never have an in-state firm listed on Pricewatch. Data on sales to such states will not help us in estimating tax sensitivities because consumers in these states can purchase from any websites on the list without paying sales tax. It would provide information on interfirm price elasticities, home-state preferences, and online-offline substitution, but the first two of these can be estimated

---

[25]Note that we do not include an "outside good" in the discrete-choice set as one might do to attempt to estimate the effect of a logit-inclusive value on aggregate demand. We are thus implicitly assuming, for example, that the total sales by Pricewatch e-retailers to state s are not affected by the states in which the e-retailers are located and the difference between the nth lowest price and the lowest price. We do this because we have little data to estimate such effects, think they must be small, and prefer a more parsimonious model in which fewer coefficients are used to capture aggregate demand effects. Reasons why any inclusive-value effects would be hard to find include that prices on Pricewatch are almost always tightly bunched, and that, in any state other than California, having more than one or two e-retailers on the list from that state is extremely rare.

precisely with sales to a smaller set of states. We decided to reduce the computational burden by carrying out our analysis on a smaller dataset containing hourly sales by our two websites in just ten states: Alabama, Florida, Georgia, Illinois, Ohio, Oregon, Pennsylvania, Texas, Virginia, and Wisconsin.

We carry out the estimation four times to obtain independent estimates using data on each of the four products: 128MB PC100 modules, 128MB PC133 modules, 256MB PC100 modules, and 256MB PC133 modules.

Note that we are assuming that it is not necessary to use instruments for the prices on the right hand side of the above equations. With regard to relative prices in the choice-between-retailers equation, we think this is a very reasonable assumption and a major reason why the Pricewatch environment is a nice one to study. We find it implausible that a substantial part of price variation is driven by information the firms have about a particular hour on a particular day being a good time to have the third as opposed to the seventh lowest price. With regard to the prices in the aggregate demand equation, one could worry more about endogeneity. These estimates are not our primary focus, however, so we are willing to think of them as coefficients on reduced-form control variables rather than as demand elasticities.

4.3   Summary Statistics

Table 4 reports summary statistics separately for each of the four types of memory modules. The unit of observation is an hour-state-website. Given that our websites sell zero memory modules to a typical state in a typical hour, average sales figures at this level are quite low. For example, the average number of 128MB PC100 modules sold by a website in one particular hour to one particular state is 0.013.[26] *Price* is the price charged by our websites. Mean prices are about $70 for 128MB modules and about $140 for the 256MB modules. The dramatic price declines that occurred over the year are visible in the minimums and maximums for this variable. *MinPrice* is the lowest price listed on Pricewatch in the hour in question. Our firm's 128MB prices are about $2 to $4 higher than this on average. Its average rank on the Pricewatch list is sixth. The average gap between our firm's 256MB

---

[26]We count a single order of multiple memory modules as having quantity one. For most of our time period, our firm limited purchases of memory modules to one per order.

price and the lowest available price is larger. Much of this is due to a period when one firm offered these modules at a very low price. Our firm's average rank is still about sixth. *PSunday* is the average value of *MinPrice* on the most recent Sunday. The summary statistics for *MinPrice − PSunday* give a feel for the within-week price volatility. This mean price difference for 128MB modules is well under a dollar but the standard deviation is large. The mean price difference for 256MB modules is -$1.39 for PC100 and -$2.07 for PC133. Again, there is a lot of variation around those means, with a minimum of -$51 and a maximum of $36 for PC133 modules! (These figures are correct.)

We do not include California in our estimation, so consumers in the ten states we are considering would not need to pay sales tax to buy from our websites. They would need to pay sales tax if they bought from an in-state firm (except for those who live in Oregon). Approximately 3% of the listed websites are in-state on average. The average tax that would be paid if buying from an in-state firm is $7.04.

4.4 Basic Results

Table 5 presents coefficient estimates obtained by performing separate nonlinear least squares estimations on the data for each of the four products: 128MB PC100, 128MB PC133, 256MB PC100, and 256MB PC133. In many ways, the four sets of results are quite similar.

The most basic fact about the Pricewatch environment is that it is intensely competitive (as we previously noted in Ellison and Ellison (2004)). The coefficients on *Price* in the four columns range from -0.47 to -0.82. The estimate for 128MB PC100 memory modules, for example, corresponds to an own-price elasticity of -33 (holding all variables fixed at their sample means). The estimates are extraordinarily significant. The decrease in demand that occurs when our firm raises its price (or is undercut) is so large as to be impossible to miss.

The coefficients on the time-trend variables illustrate the growth (and decline) of Pricewatch over our sample period. The coefficient on *TimeTrend1* in the first column indicates that that overall demand was growing at about 2% per day (equivalent to 60% per month) in the first three months of our sample (May-August, 2000). Growth rates for later periods are obtained by adding all of the earlier coefficients. This suggests that sales decreased 40%

per month in the fall of 2000, were flat in the winter 2000-01, and fell an additional 20% per month in the spring of 2001. Growth rates for the other three products are similar, suggesting these patterns are not just product-specific fluctuations.

Figure 3 presents a graph of the hour dummies.[27] They indicate that online shopping picks up substantially between 7am and 11am, continues at approximately the 11am level past the normal workday, remains at about 80% of peak value until midnight, and then drops off substantially until 6am. The large number of late-night purchases suggests that greater availability may be an important factor differentiating e-retail from traditional retail.

Intraday Sales Pattern:
128MB PC100 Modules
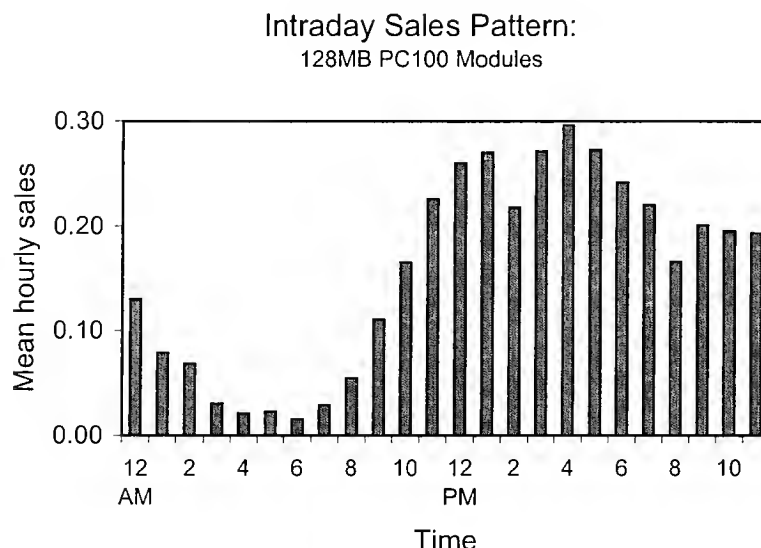
Figure 3: Intraday Sales Pattern: 128MB PC100 memory modules

4.5 Taxes

Recall that in our demand specification consumers are assumed to evaluate products on the basis of $Price + \beta_2 SalesTax$, with $SalesTax$ measured in dollars. Hence, an estimate of one on the $SalesTax$ coefficient would correspond to the standard rational model in which consumers care only about their total expenditure and an estimate of zero would

---

[27]Recall that we simply set these to the sample mean quantities for each hour rather than making them part of the nonlinear least squares estimation. Sample means are computed on a time-zone adjusted basis with the times of all purchases being recorded from the consumer's perspective.

correspond to consumers who are entirely insensitive to tax differences.[28]

The most general conclusion we draw from the four sets of results is that consumers pay less attention to sales taxes than the standard rational model predicts. The estimates in the four columns are 0.05 (s.e. 0.12), 0.38 (s.e. 0.15), 0.10 (s.e. 0.09), and 1.27 (s.e. 0.64). Note that the first three are significantly different from unity while the second and last are significantly different from zero. We interpret this as evidence that consumers are paying attention to taxes but not as much as price differences of a similar magnitude. These effects are not very precisely estimated, though.

It is important to note that the fact that consumers pay less attention to tax differences than to price differences does not imply that sales taxes are not important. Our consumers are extraordinarily sensitive to price differences, so even if the coefficient on the *SalesTax* variable was 0.3, our estimates would be that a firm that must collect a 6% sales tax would have its sales decline by about 60%.

4.6 Geography

Geography enters our demand model in two ways. First, *ShippingTime* allows for the possiblity that consumers may prefer to buy from e-retailers in nearby states because they will have faster delivery times with standard ground shipping. We fail to find evidence of such an effect in these regressions: only two of the four estimates are negative; only one of these is significant. If one thinks about the magnitudes of these estimates relative to the price coefficients, one would conclude that any geographic effects are small. A coefficient of 0.1 on the *ShippingTime* variable would mean that reducing the shipping time by one day is comparable to reducing the price by about 20 cents. Recall that our cross sectional regressions provided some evidence of a shipping time effect, in contrast to what we find here.

Second, we included the *HomeState* dummy to allow for the possibility that consumers may have an additional preference for buying from in-state firms. Here, we get consistent

---

[28]There are clearly other "rational" models in which the coefficient would be greater than or less than one. An example of the former is if price is a signal of quality so that a high price-zero tax offer is preferable to a low price-high tax offer with the same total expenditure. Examples of the latter would be a model in which consumers benefit from taxes collected by their state government or have nonselfish preferences and thereby gain from payments to local firms and/or governments.

results showing that geography does matter. Three of the four coefficient estimates are positive and significant. The magnitudes of the coefficient estimates indicate that the home-state preference will roughly offset a two dollar price difference. In light of our earlier estimates that consumers pay less attention to difference in sales taxes than to differences in prices, it also implies that the home-state preference will outweigh the sales-tax disadvantage on moderately priced items. For example, if the $SalesTax$ coefficient is 0.33, the $Price$ coefficient is -0.5, the $HomeState$ coefficient is 1.0 and the tax rate is 6%, then the home-state preference will outweigh the tax disadvantage on items costing $100 or less. The finding that the home state preference is nearly strong enough to outweigh the tax-disadvantage of buying from an in-state firm contrasts with our earlier finding that our firm sells much less in California than in other states. It is possible that some states enjoy home state preference while others do not.

4.7   Online-Offline Substitution

As noted above, the within-week variation in the online price provides an opportunity to examine online-offline substitution: when offline prices are constant over the course of a week, the online-offline price gap will move one-for-one with changes in the online price. In terms of the variables defined above, the within-week change in the lowest price available on Pricewatch is $MinPrice - PSunday$. Our specification includes these two variables separately in the equation for the number of online consumers.

The $MinPrice$ variable is the lowest price available on Pricewatch in the hour in question. Note that it may affect the number of consumers buying memory on Pricewatch for two reasons: aggregate demand will be higher when the price is lower; and a higher share of consumers will buy online (as opposed of offline) when the online-offline price gap is wider. Our estimates of this coefficient are highly significant and consistent across the four product classes: a one-dollar decrease in the price of a memory module is estimated to increase the total demand at Pricewatch retailers by about 3%. This corresponds with an elasticity about -2 for the 128MB modules and about -4 for the 256MB modules (at the mean prices). We take this result as suggestive of substantial online-offline substitution, because it seems unlikely that the aggregate demand for memory modules is so elastic. One piece of evidence

to this effect is that collusion between DRAM manufacturers led to a rougly 400% increase in memory prices between December 2001 and May of 2002.[29] Optimal collusion would result in a much smaller price increase if aggregate demand was so elastic.

PSunday is the lowest price listed on Pricewatch on the most recent Sunday. One would expect higher values of this variable to be associated with higher sales through Pricewatch for two reasons: it should be associated with a higher offline price; and the pool of consumers interested in buying memory at the current price may be larger when past prices were higher. Our estimates on the 128MB modules provide little evidence of either effect. *PSunday* is significant in two of the four estimations, but one of the significant estimates is negative. Among the potential explanations for our inability to find the predicted effects are that our data are limited, that consumers may not be reacting strongly to transient differences in online-offline prices because they are unaware of them, and that there may not be much of a customer-pool effect because few consumers consider purchasing at multiple points in time.

5   Conclusion

In this paper we have examined Internet retail demand using two different approaches: a cross-sectional analysis of demand in different states and a discrete-choice analysis of demand at an hourly frequency. The two analyses exploit separate sources of variation in the data: the state-level analysis ignores all of the variation in competitive conditions; and the discrete-choice analysis uses state fixed effects to absorb any persistent factors like tax rates.

Our most basic conclusion on sales taxes is that they are an important driver of e-retail activity. Our state-level regressions show clearly that sales are higher in states that levy higher sales taxes on traditional retail purchases. The fact that the websites we study sell so little in California is strong evidence that what we are picking up is a tax effect and not some artifact of unobserved heterogeneity. The environment we study is somewhat unusual in that consumers are highly savvy and price-sensitive, but in this environment at least,

---

[29]The largest manufacturers reached agreements with the DOJ in 2004 to 2006. These included $700 million in fines and jail time for senior executives at Infineon, Hynix, and Samsung.

we would agree with Goolsbee's (2001) conclusion that applying sales taxes to e-retail sales could reduce e-retail demand by one-quarter or more. Our discrete-choice analysis is not inconsistent with this conclusion, but is a little surprising in light of our earlier results. We find that consumers do not pay as much attention to differences in taxes as they do to differences in pre-tax prices when choosing between e-retailers. Taxes do matter to consumers, though, and given how tightly distributed prices in this market are, they can have large effects on consumer behavior.

The state-level analysis indicates that geography still matters in e-retail. The websites we study make more sales to states that are closer to California in a shipping-time metric. We do not find an analogous effect, however, in our discrete-choice analysis. One thing we do find consistently in the discrete-choice analysis is that consumers have a preference for buying from in-state e-retailers. We think this is an interesting result on the sources of geographic differentiation. It has implications for market structure that would differ from what one would obtain from thinking about shipping times. A world where consumers care about purchasing from their home state could lead to a less concentrated e-retail sector with many small firms, whereas a world where consumers do not have a home-state preference but do care about shipping times could lead to a sector dominated by a few large firms that effectively use distributed warehouses to minimize both shipping times and sales tax liabilitites.

A couple of our results suggest that there is substantial substitution between online and offline retail. Most states have few e-retailers listed on Pricewatch. Accordingly, most of the tax-demand relationship we identify in the state-level analysis must be coming from substitution away from traditional retail. Similarly, we think of our estimate of the effect of prices on the total Pricewatch demand as being too large to be due to market expansion effects and hence must be coming at least in part from online-offline substitution. We do not, however, find convincing evidence of short-term online-offline substitution from our analysis of recent past prices.

Taken together we see our results as suggesting that there are limitations to consumer rationality. In each of the dimensions we analyze, there are some considerations that are quite easy for consumers to recognize and some that are more subtle or would entail more

costly information aquisition. A general theme that appears to emerge from our analysis is that consumers are closer to the traditional rational ideal on the easier tasks.

In terms of sales taxes, for example, it is quite easy for consumers, in high tax states, especially, to learn the general principle that buying things online saves on taxes, and we find clear evidence sales being higher in high tax states. That the tax benefit will be larger on more expensive items is a little more subtle, and the evidence on whether tax rate-sensitivity is larger for more expensive items in our data is less clear. Our discrete-choice analysis also indicates that consumers do not make a one-for-one tradeoff between differences in item prices and sales taxes. It is intuitive that the way that data is presented on Pricewatch might lead to such a result if consumers are boundedly rational. Pricewatch's lists includes the state in which each firm is located so that a consumer who understood that taxes would only be assessed on in-state purchases (and knew his local sales tax rate) could compute the sales taxes, but taxes are not presented to the consumer and the lists are always sorted on the basis of the pre-tax price. In this vein it is noteworthy that Brynjolfsson and Smith's (2001) contrasing estimates—they find that react twice as strongly to tax differences as they do to item price differences—were obtained in an environment in which taxes are explicity presented to consumers in a list that is sorted on the basis of tax-inclusive prices.[30]

Similar patterns exist in the other dimensions of behavior we consider. The tax differences we study are stable over time. This should make it very easy for consumers in high tax states to learn that purchasing online is a good idea. The high-frequency volatility of memory module prices on Pricewatch is highly unusual. We find it plausible that we may have a harder time finding consumer reactions to these short run price movements because (a) "rational" consumers with information acquisition costs may not have invested in up-to-date infomation on the evolution of the online-offline price gap over the last few days, or (b) "boundedly rational" consumers may not have figured out that they can exploit retail stores' "stale" prices when there has been a very recent run up in online prices.

The environment we study is somewhat special. The consumers shopping through

---

[30]Hossain and Morgan (2006) find that consumers do not fully take shipping costs into account in a neatly-designed field experiment involving selling items on eBay. A commonality between shipping costs in their experiment and tax differences on Pricewatch is that the shipping cost differences were easily available in the item descriptions, but some effort would have been required to learn the differences.

Pricewatch are more technically savvy than most and we presume much more price sensitive. They are presented with more complete data on competing price offers than are usually available. In other dimensions, however, the environment might not be so unrepresentative. The fact that Pricewatch does not compute or highlight differences in sales taxes and the fact that it lacks information on offline prices may lead to tax- and online-offline substitution effects that are not so different from those we would see in other environments. Limitations on consumer rationality may also make it easier to generalize from our results: it is much easier to generalize in a model where a fraction of consumers always buy online to save on taxes than in a model where behavior depends in a complex way on the number of dollars in taxes that can be avoided, for instance.

Technically, our analysis is standard. What could perhaps be more broadly useful is our suggestion that discrete-choice models may be usefully applied to datasets containing quantity data for one firm. Price data for all of the firms in a market are fairly easy to come buy. Quantity data are much harder to obtain. There may, however, be many other situations like ours where quantity data could be obtained from one firm. (This could even be done in a field experiment.) Our example suggests that this may be a fruitful way to explore interfirm competition.

References

Baye, Michael, Rupert J. Gatti, Paul Kattuman, and John Morgan (2005). "Estimating Firm-Level Demand at a Price Comparison Site: Accounting for Shoppers and the Number of Competitors", University of California-Berkeley, mimeo.

Baye, Michael, and John Morgan (2004). "Price Dispersion in the Lab and on the Internet: Theory and Evidence," *Rand Journal of Economics*, 35 (3), 449-466.

Brown, Jeffrey and Austan Goolsbee (2002). "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy*, 110 (3), 481-507.

Brynjolfsson, Erik and Michael Smith (2001). "Consumer Decision-making at an Internet Shopbot," *Journal of Industrial Economics*, 49 (4), 541-558.

Chevalier, Judith and Austan Goolsbee (2003). "Price Competition Online: Amazon Versus Barnes And Noble," *Quantitative Marketing and Economics*, 1 (2), 203-222.

Chiou, Leslie (2005). "Empirical Analysis of Retail Competition: Spatial Differentiation at Walmart, Amazon.com, and Their Competitors," Occidental College, mimeo.

Davis, Peter (2006). "Spatial Competition in Retail Markets: Movie Theaters," *RAND Journal of Economics*, forthcoming.

Ellison, Glenn (2005). "A Model of Add-on Pricing," *Quarterly Journal of Economics*, 120 (2), 585-637.

Ellison, Glenn and Sara Fisher Ellison (2004). "Search, Obfuscation, and Price Elasticities on the Internet," NBER Working Paper 10570.

Ellison, Glenn and Sara Fisher Ellison (2005). "Lessons about Markets from the Internet," *Journal of Economic Perspectives*, 19 (2), 139-158.

Fox, William (1986). "Tax Structure and the Location of Economic Activity Along State Border," *National Tax Journal*, 39 (4), 387-401.

Goolsbee, Austan (2000). "In a World without Borders: The Impact of Taxes on Internet Commerce," *Quarterly Journal of Economics*, 115 (2), 561-576.

Goolsbee, Austan (2001). "Competition in the Computer Industry: Online Versus Retail," *Journal of Industrial Economics*, 49 (4), 487-499.

Greene, William H. (1997): *Econometric Analysis*. Prentice Hall.

Hausman, Jerry, Bronwyn Hall and Zvi Griliches (1984): "Economic Models for Count Data with an Application to the Patents-R&D Relationip," *Econometrica* 52, 909-938.

Hossain, Tanjim and John Morgan (2006): "...Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay," *Advances in Economic Analysis & Policy*,

6, Article 3.

Prince, Jeffrey (2005): "Substitutability between Online and Retail Personal Computers: Changes over Time and the 'Online Research' Component," mimeo, Cornell University.

Walsh, Michael and Jonathan Jones (1988). "More Evidence on the 'Border Tax' Effect: The Case of West Virginia," *National Tax Journal*, 41 (2), 261-265.

Weisbrod, G., R. Parcells and C. Kern (1984). "A Disaggregate Model for Predicting Shopping Area Market Attraction," *Journal of Retailing*, 60 (1), 65-83.

Variable	Mean	St.Dev	Min	Max
$Quantity128$	203.5	176.0	19.0	762.0
$Quantity256$	85.6	84.3	5.0	391.0
$OfflineSalesTaxRate$	0.057	0.021	0.000	0.084
$InternetAccess$	0.57	0.07	0.41	0.70
$ShippingTime$	3.89	0.92	1.50	5.00
$\frac{ComputerStores}{GasStations}$	0.092	0.034	0.041	0.184
$\log(Population)$	15.02	1.04	13.11	17.33

Table 1: Summary statistics for state-level regressions

	Product/Estimation Method			
	128MB	256MB	128MB	256MB
$Offline Sales Tax Rate$	5.96	6.33	6.08	7.47
	(2.16)	(2.59)	(2.39)	(2.71)
$California$	-1.03	-0.84	-1.01	-0.85
	(4.01)	(3.21)	(3.18)	(2.46)
$Shipping Time$	-0.10	-0.07	-0.09	-0.05
	(2.04)	(1.32)	(1.64)	(0.82)
$Internet\ Access$	1.89	1.04	1.86	0.91
	(2.62)	(1.25)	(2.32)	(1.06)
$\frac{Computer Stores}{Gas Stations}$	1.90	4.39	1.94	5.05
	(1.11)	(2.29)	(0.98)	(2.36)
$\log(Population)$	0.85	0.89	0.87	0.91
	(20.54)	(18.98)	(18.51)	(17.88)
Estimation	Neg. Binomial		OLS	OLS
Observations	51	51	51	51
R^2			0.93	0.93

Note: The first two columns report estimates from negative binomial regressions with *Quantity128* and *Quantity256* as dependent variables. The third and fourth columns present estimates from related OLS regressions with the log(*Quantity*) as the dependent variable. *t*-statistics in parentheses.

Table 2: State-level regressions

	Product/Sample Period			
Product:	128MB		256MB	
Price-based sample	$20-$50	$100+	$20-$50	$100+
$Sales Tax Rate$	4.22	7.32	7.43	8.89
	(1.74)	(2.95)	(2.26)	(2.62)
$California$	-0.92	-1.39	-0.62	-0.74
	(3.26)	(6.29)	(2.24)	(2.37)
$Ship Time$	-0.10	-0.15	-0.00	0.01
	(1.88)	(2.79)	(0.01)	(0.13)
$Internet Access$	1.13	3.67	-0.69	1.97
	(1.38)	(4.32)	(0.62)	(1.63)
$\frac{Computer Stores}{Gas Stations}$	1.05	1.61	5.59	6.25
	(0.55)	(0.88)	(2.24)	(2.38)
$Log(Population)$	0.85	0.94	0.80	0.85
	(18.05)	(19.26)	(13.20)	(12.91)

Note: The table reports estimates from negative binomial regressions. The dependent variable is the number of orders received from each state during the time period when the lowest price listed on Pricewatch was in the specified range. *t*-statistics in parentheses.

Table 3: State-level regressions examining sales in different time periods

Variable	Mean	St.Dev	Min	Max	Mean	St.Dev	Min	Max
	128MB PC100				128MB PC133			
Quantity	0.013	0.118	0	4	0.010	0.105	0	2
Price	65.52	34.56	21	123	74.11	36.70	21	131
MinPrice	61.71	33.46	20	122	71.72	37.27	20	131
Rank	6.5	4.1	1	21	6.0	4.5	1	21
MinPrice − PSunday	-0.24	3.30	-13	13	-0.33	3.54	-16	16
	Number of Obs.: 154070				Num. Obs.: 133310			
	256MB PC100				256MB PC133			
Quantity	0.004	0.068	0	3	0.007	0.091	0	4
Price	129.98	65.10	43	258	146.17	79.27	39	291
MinPrice	120.19	58.29	43	215	135.24	75.54	39	269
Rank	5.9	3.1	1	12	6.3	3.05	1	12
MinPrice − PSunday	-1.39	5.12	-18.75	35	-2.07	8.21	-51	36
	Number of Obs.: 120420				Num. Obs.: 124520			

Table 4: Summary statistics for individual-level regressions

	Product			
	128MB PC100	128MB PC133	256MB PC100	256MB PC133
Variables affecting choices between sites				
Price	-0.53	-0.82	-0.49	-0.47
	(33.16)	(28.64)	(20.39)	(28.37)
SalesTax	0.05	0.38	0.10	1.27
	(0.37)	(2.52)	(1.07)	(1.98)
HomeState	1.24	1.31	1.13	1.23
	(3.69)	(2.92)	(2.31)	(1.19)
ShippingTime	0.14	-0.02	-0.06	-0.11
	(2.34)	(0.30)	(0.74)	(2.01)
SecondScreen	-1.58	0.61		
	(1.05)	(0.18)		
Variables affecting total Pricewatch demand				
Weekend	-0.43	-0.34	-0.31	-0.80
	(10.88)	(7.36)	(4.11)	(11.41)
MinPrice	-0.03	-0.03	0.00	-0.03
	(5.51)	(4.47)	(0.19)	(9.27)
PSunday	0.001	0.003	-0.033	0.015
	(0.17)	(0.49)	(4.67)	(3.29)
TimeTrend1	0.02	0.02	0.02	0.02
	(5.29)	(6.93)	(4.71)	(3.37)
TimeTrend2	-0.03	-0.02	-0.03	-0.04
	(4.89)	(3.98)	(4.25)	(3.68)
TimeTrend3	0.02	0.00	0.00	0.01
	(4.78)	(0.17)	(0.53)	(1.50)
TimeTrend4	-0.01	-0.01	0.01	0.01
	(5.84)	(2.85)	(6.36)	(8.23)
Observations	154070	133310	120420	124520
R^2	0.04	0.03	0.02	0.04

Note: Dependent variables are number of distinct customers in each of ten states ordering from each of websites A and B in each of approximately 7900 hours. Regressions also contain state and website dummies. t-statistics are in parentheses.

Table 5: Discrete-choice model of hourly sales of memory modules in ten states